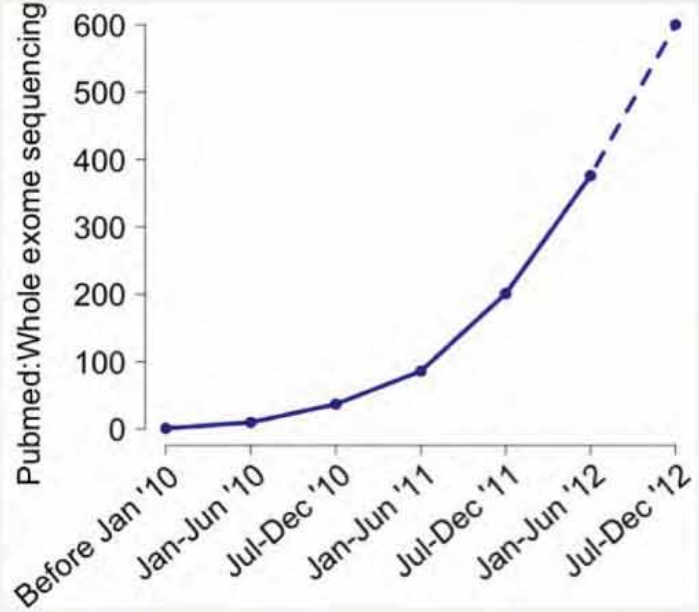
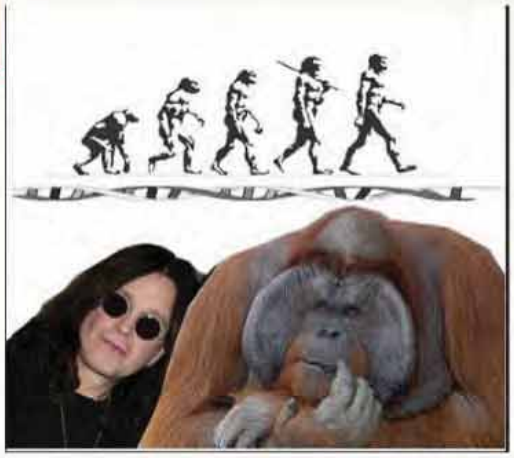
Insights: Identification of Candidate Variants Using Exome Data in Ophthalmic Genetics

Agenda

- Landscape of NGS
- Overview of our research
- Using Exome data to filter-
  - Step by step application of techniques (Tips and Tricks)
    - Filter by Marker Statistics
    - Filter by Gene List
    - Filter by Functional Predictions
    - Filter by Multiple Columns
- Challenges/Opportunities



Gerald Goh, Genomics Inform 2012;10(4):214-219



Terri Young, MD, MBA

Professor of Ophthalmology, Pediatrics and Medicine

-Duke University Center for Human Genetics (CHG)
-Duke Eye Center
-Duke-National University of Singapore -Graduate Medical School

Our research interests

nature.com » journal home » current issue » letter » full text

NATURE GENETICS | LETTER

Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia

Virginie J M Verhoeven, Pirro G Hysi, Robert Wojciechowski, Qiao Fan, Jeremy A Guggenheim, René Höhn, Stuart MacGregor, Alex W Hewitt, Abhishek Nag, Ching-Yu Cheng, Ekaterina Yonova-Doing, Xin Zhou, M Kamran Ikram, Gabrielle H S Buitendijk, George McMahon, John P Kemp, Beate St Pourcain, Claire L Simpson, Kari-Matti Mäkelä, Terho Lehtimäki, Mika Kähönen, Andrew D Paterson, S Mohsen Hosseini, Hoi Suen Wong, Yanhui Xu + et al.

Affiliations | Contributions | Corresponding author

Nature Genetics **45**, 314–318 (2013) | doi:10.1038/ng.2554

Received 03 October 2012 | Accepted 16 January 2013 | Published online [illegible] February 2013

OPEN ACCESS Freely available online PLOS GENETICS

Genetic Variants on Chromosome 1q41 Influence Ocular Axial Length and High Myopia

Qiao Fan[1], Veluchamy A. Barathi[2,3], Ching-Yu Cheng[1,2,3], Xin Zhou[1], Akira Meguro[4], Isao Nakata[5,6], Chiea-Chuen Khor[7,8,9], Liang-Kee Goh[7,10,11], Yi-Ju Li[12,13], Wan'e Lim[2], Candice E. H. Ho[2], Felicia Hawthorne[14], Yingfeng Zheng[2], Daniel Chua[2], Hidetoshi Inoko[15], Kenji Yamashiro[4], Kyoko Ohno-Matsui[12], Keitaro Matsuo[16], Fumihiko Matsuda[6], Eranga Vithana[2,3], Mark Seielstad[17], Nobuhisa Mizuki[4], Roger W. Beuerman[2,4,18], E-Shyong Tai[1,19], Nagahisa Yoshimura[5], Tin Aung[2,4], Terri L. Young[14,11], Tien-Yin Wong[1,2,4,19], Yik-Ying Teo[1,7,20,21,*], Seang-Mei Saw[1,2,4,19,20,*]

ORIGINAL ARTICLE

Whole exome sequencing identifies a mutation for a novel form of corneal intraepithelial dyskeratosis

Vincent José Soler,[1,2] Khanh-Nhat Tran-Viet,[1] Stéphane D Gallacy,[2] Vachiranee Limviphuvadh,[3] Thomas Patrick Klemm,[4] Elizabeth St Germain,[1] Pierre R Fournié,[2,3] Céline Guillaud,[2,3] Sebastian Maurer-Stroh,[3,6] Felicia Hawthorne,[1] Cyrielle Suarez,[2,3] Bernadette Kantelip,[2] Natalie A Afshari,[8] Isabelle Creveaux,[9] Xiaoyan Luo,[1] Weihua Meng,[2] Patrick Calvas,[2] Myriam Cassagne,[2,3] Jean-Louis Arné,[3] Steven G Rozen,[4] François Malecaze,[2,3] Terri L Young[1,8]

was initially reported in a Native American

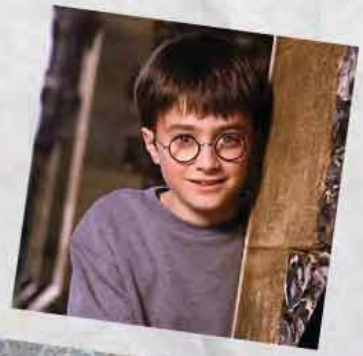
Hum Genet (2012) 131:467–480
DOI 10.1007/s00439-012-1264-0

ORIGINAL INVESTIGATION

Large scale international replication and meta-analysis st confirms association of the 15q14 locus with myopia. The CREAM consortium

Virginie J. M. Verhoeven · Pirro G. Hysi · Seang-Mei Saw · Veronique Vitart · Alireza Mirsh Jeremy A. Guggenheim · Mary Frances Cotch · Kenji Yamashiro · Paul N. Baird · David A. M Robert Wojciechowski · M. Kamran Ikram · Alex W. Hewitt · Priya Duggal · Sarayut Janmal Chiea-Chuen Khor · Qiao Fan · Xin Zhou · Terri L. Young · E-Shyong Tai · Liang-Kee Goh · Yi-Ju Li · Tin Aung · Eranga Vithana · Yik-Ying Teo · Wanting Tay · Xueling Sim · Igor Rudan Caroline Hayward · Alan F. Wright · Ozren Polasek · Harry Campbell · James F. Wilson · Bria Isao Nakata · Nagahisa Yoshimura · Ryo Yamada · Fumihiko Matsuda · Kyoko Ohno-Matsui ·

- Myopia (nearsightedness)
- Primary Congenital Glaucoma
- Stickler/Wagner Syndromes
- Corneal Dystrophy
- Strabismus (crossed eye)
- Eyelid Malformation
- Microphthalmia/Anopthalmia



Why Vision Research?

Doctor visits,contacts, glasses, etc --> $200 annually in US

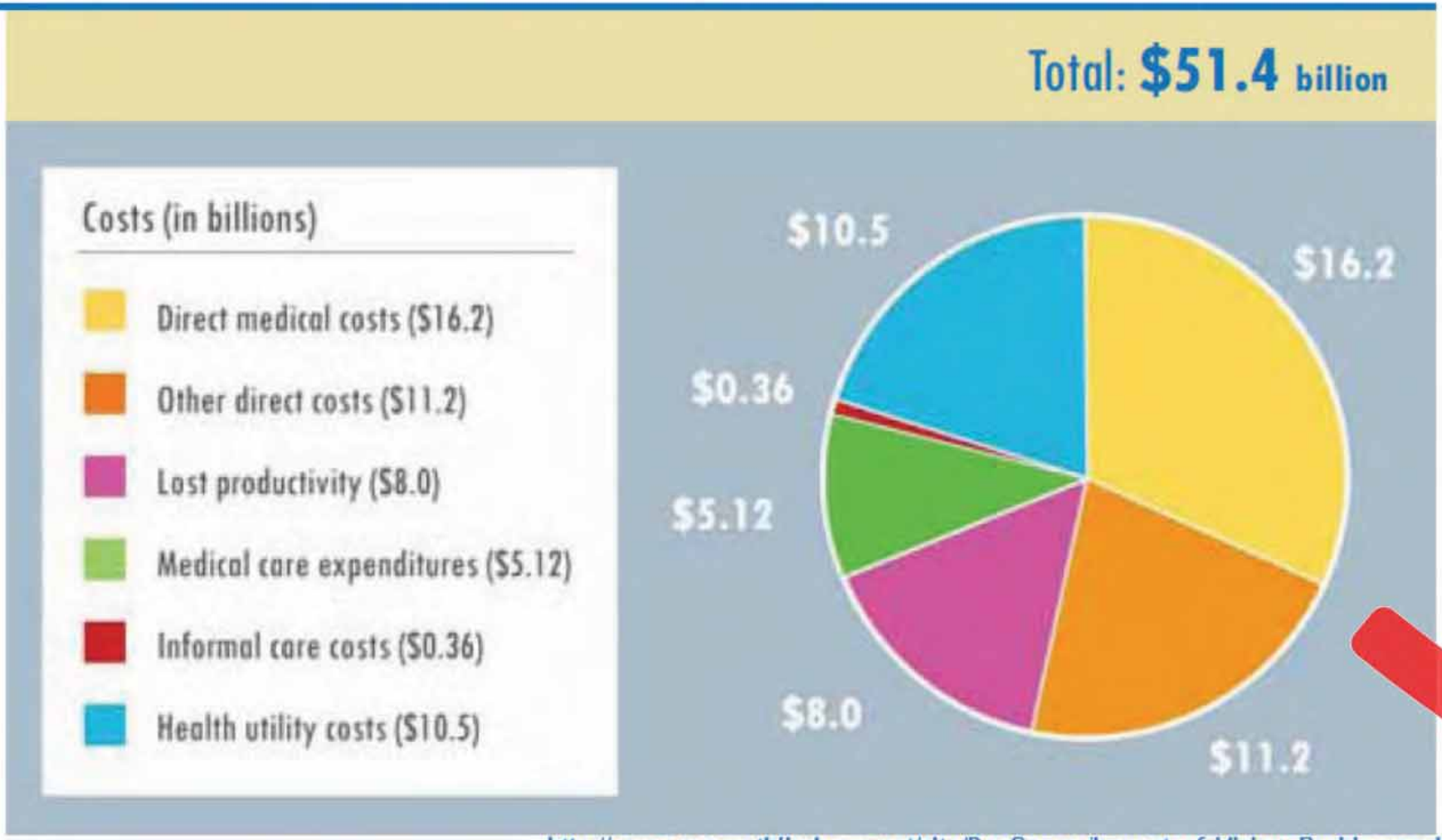
Myopia related costs exceeds $14 billion dollars annually

Causes

- Genetic
  - Twin Studies
  - Animal studies
- Environmental
  - Epidemiology studies

Economic Impact

Graph 1.1 Total Annual Economic Impact of Vision Problems in the U.S.



http://www.preventblindness.net/site/DocServer/Impact_of_Vision_Problems.pdf

Doctor visits,contacts, glasses, etc --> $200 annually in US

Myopia related costs exceeds $14 billion dollars annually

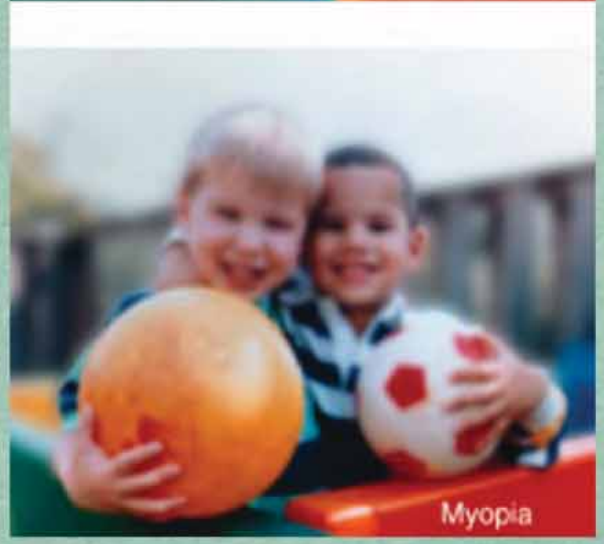
Children

The most common types seen in children are:

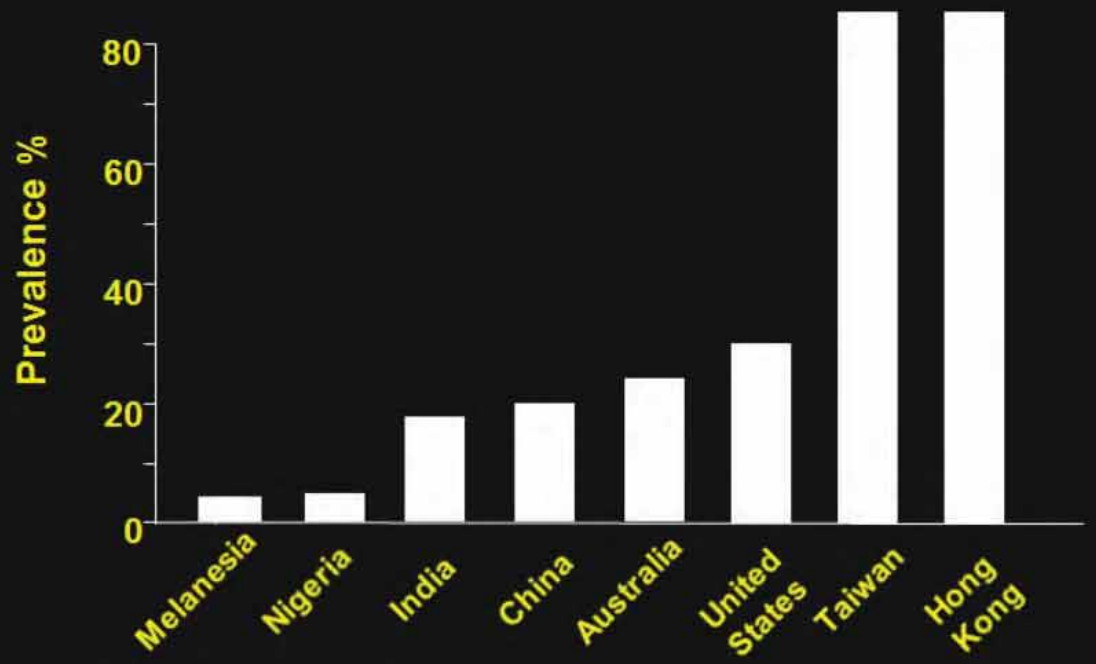
Myopia (nearsightedness)

Strabismus (crossed eyes)

Amblyopia (lazy eye)



Prevalence of myopia



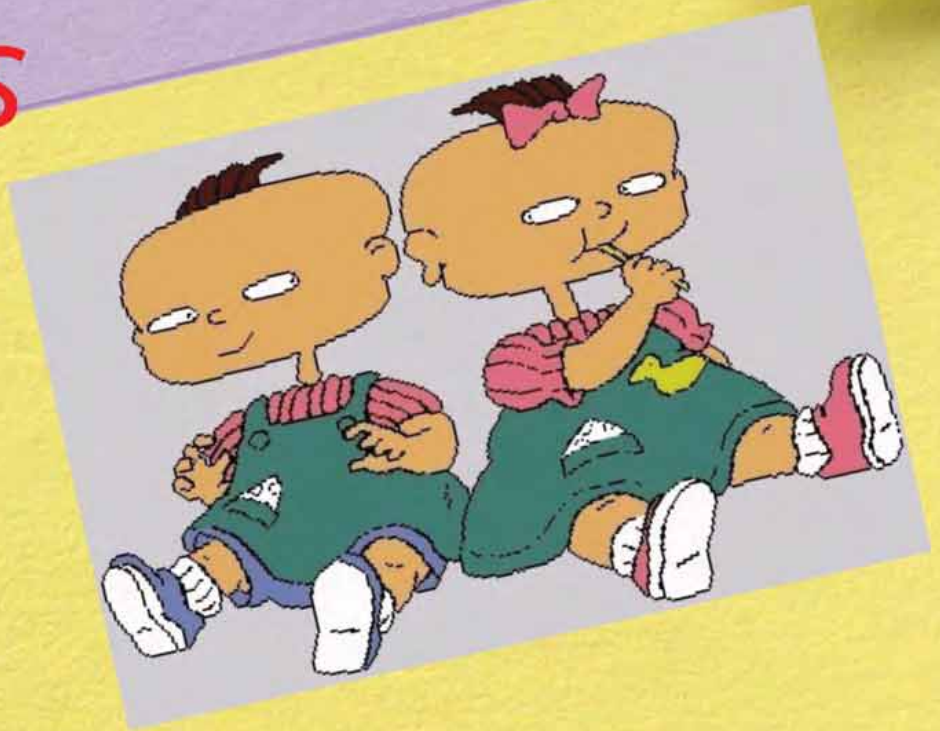
- 19.45% - India (Dandona et al. 2002)
- 21.8% - China (Xu et al 2005)
- 33.1% - US (Vitale et al. 2008)
- 84% - Taiwan (Lin et al. 2004)
- 85% - Hong Kong (Lam et al. 2004)

- 3% - Prevalence of high myopia (≤-6D) (Vongphanit et al. 2002)

- 19.45% - India (Dandona et al. 2002)
- 21.8% - China (Xu et al 2005)
- 33.1% - US (Vitale et al. 2008)
- 84% - Taiwan (Lin et al. 2004)
- 85% - Hong Kong (Lam et al. 2004)

- 3% - Prevalence myopia (≤-6D) (Vongphanit et al. 2002)

Causes



- Genetic
  - Twin Studies
  - Animal studies
- Environmental
  - Epidemiology studies

Stickler Syndrome

Clinicallly variable and heterogenous disorder

Prevalence of 1 in 10,000

-Ocular
-Auditory
-Skeletal
-Orofacial

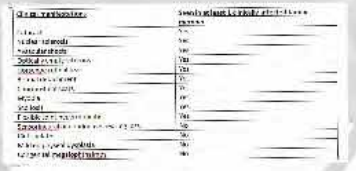
Genes

Gene Symbol	% of Disease Attributed to Mutations in This Gene	Inheritance
COL2A1	80%-90%	Autosomal Dominant
COL11A1	10%-20%	Autosomal Dominant
COL11A2	Rare, unknown	Autosomal Dominant
COL9A1	Rare, unknown	Autosomal Recessive
COL9A2	Rare, unknown	Autosomal Recessive

*GeneReviews

Family

- 49 members (16 affecteds)
- Initially diagnosed Wagner Syndrome
- Variable clinical presentations
- Screened for known mutations



Filtering Assumptions

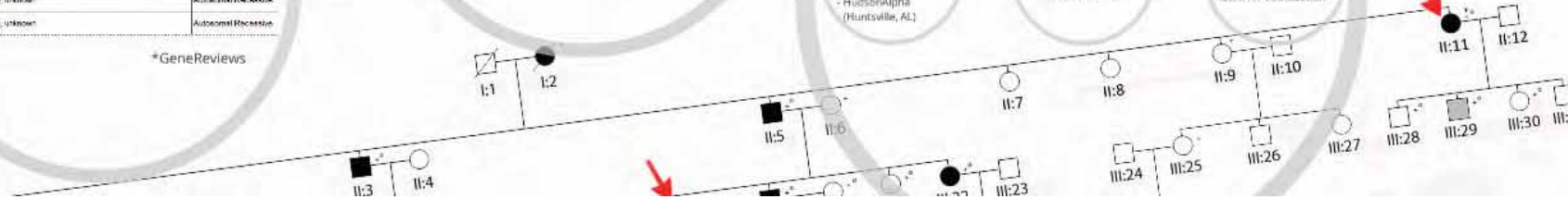
- Rare
- Autosomal Dominant model
- Coding-Nonsynonymous
- Conserved
- Biologically relevant

Processes

Input
- 2 affecteds for Exome
- Nimblegen V2 Lib Capture (50Mb)
- HiSeq1000
- HudsonAlpha (Huntsville, AL)

Bioinformatics
- Aligned, removed duplicates, GATK to call variants

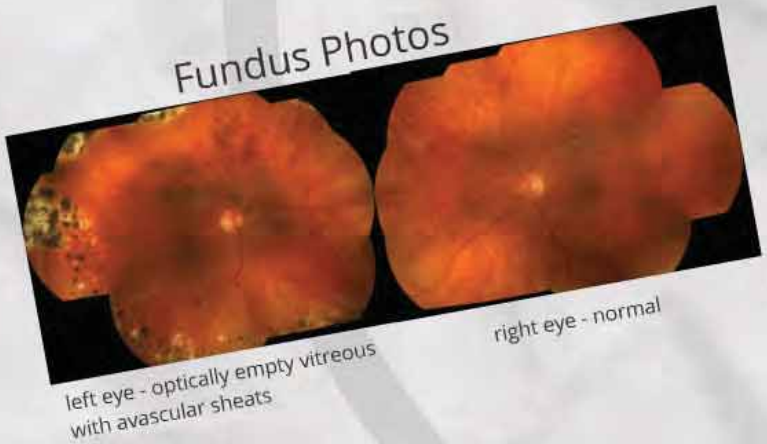
Output
* VCF (Variant Calling File) for analysis
* Bam for visualization



Stickler Syndrome

Clinicallly variable and heterogenous disorder

Prevalence of 1 in 10,000



-Ocular
-Auditory
-Skeletal
-Orofacial

Genes

Gene Symbol	% of Disease Attributed to Mutations in This Gene	Inheritance
COL2A1	80%-90%	Autosomal Dominant
COL11A1	10%-20%	Autosomal Dominant
COL11A2	Rare, unknown	Autosomal Dominant
COL9A1	Rare, unknown	Autosomal Recessive
COL9A2	Rare, unknown	Autosomal Recessive

*GeneReviews

Family

- 49 members (16 affecteds)
- Initially diagnosed Wagner Syndrome
- Variable clinical presentations
- Screened for known mutations

Clinical manifestations	Seen in at least 1 clinically affected family member
Cataract	Yes
Nuclear sclerosis	Yes
Avascular sheets	Yes
Optically empty vitreous	Yes
Horsehoe retinal tear	Yes
Retinal detachment	Yes
Chorioretinal scars	Yes
Myopia	Yes
Scoliosis	Yes
Flexible Joint hypermobility	Yes
Sensorineural or conductive hearing loss	No
Cleft palate	No
Mild epiphyseal dysplasia	No
Congenital megalophthalmos	No

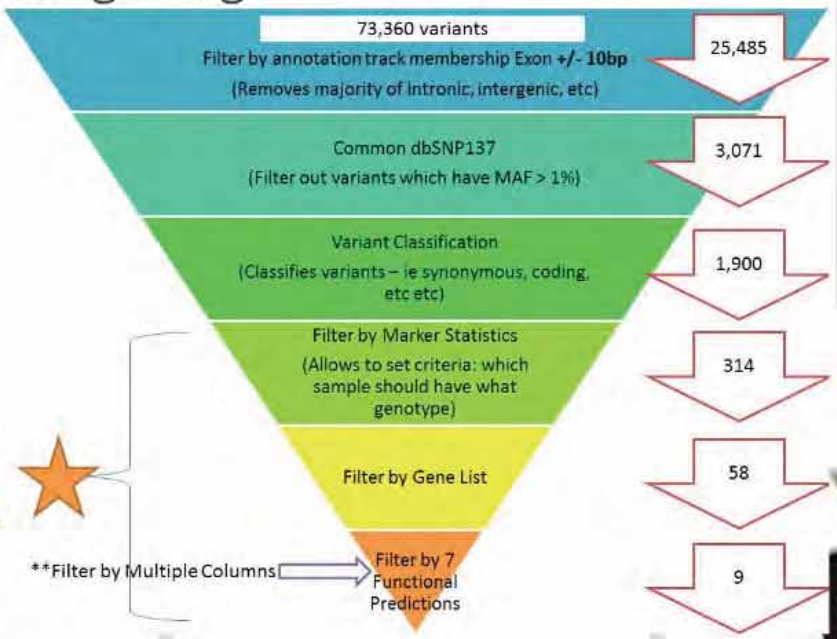
- 2 a
- Ni
  Ca
- Hi
- Hu
  (Hu

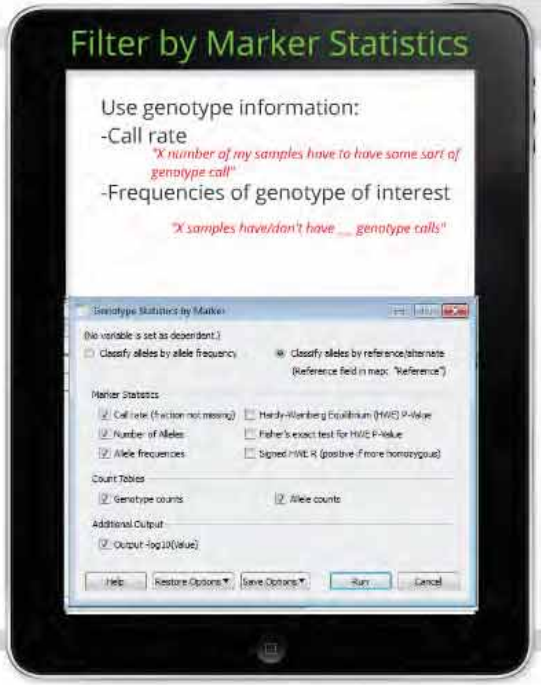
ily
ecteds)
Wagner Syndrome
sentations
mutations
• Biologically relevant
Seen in at least 1 clinically affected family member
Yes
Yes
Yes
Yes
Yes
Yes
Yes
Yes
Yes
Yes
No
No
No
No
Processes
Input
• 2 affecteds for Exome
• Nimblegen V2 Lib Capture (50Mb)
• HiSeq1000
• HudsonAlpha (Huntsville, AL)
Bioinformatics
• Aligned, removed duplicates, GATK to call variants
Output
*.VCF (Variant Calling File) for analysis
*.Bam for visualization
II:5
II:6
II:7
II:8
II:9
II:10
II:11
III:17
III:18
III:19
III:20
III:21
III:22
III:23
III:24
III:25
III:26
III:27
III:28
III:29
IV:17
IV:18
IV:19
IV:20
IV:21
IV:22
IV:23
IV:24
IV:25

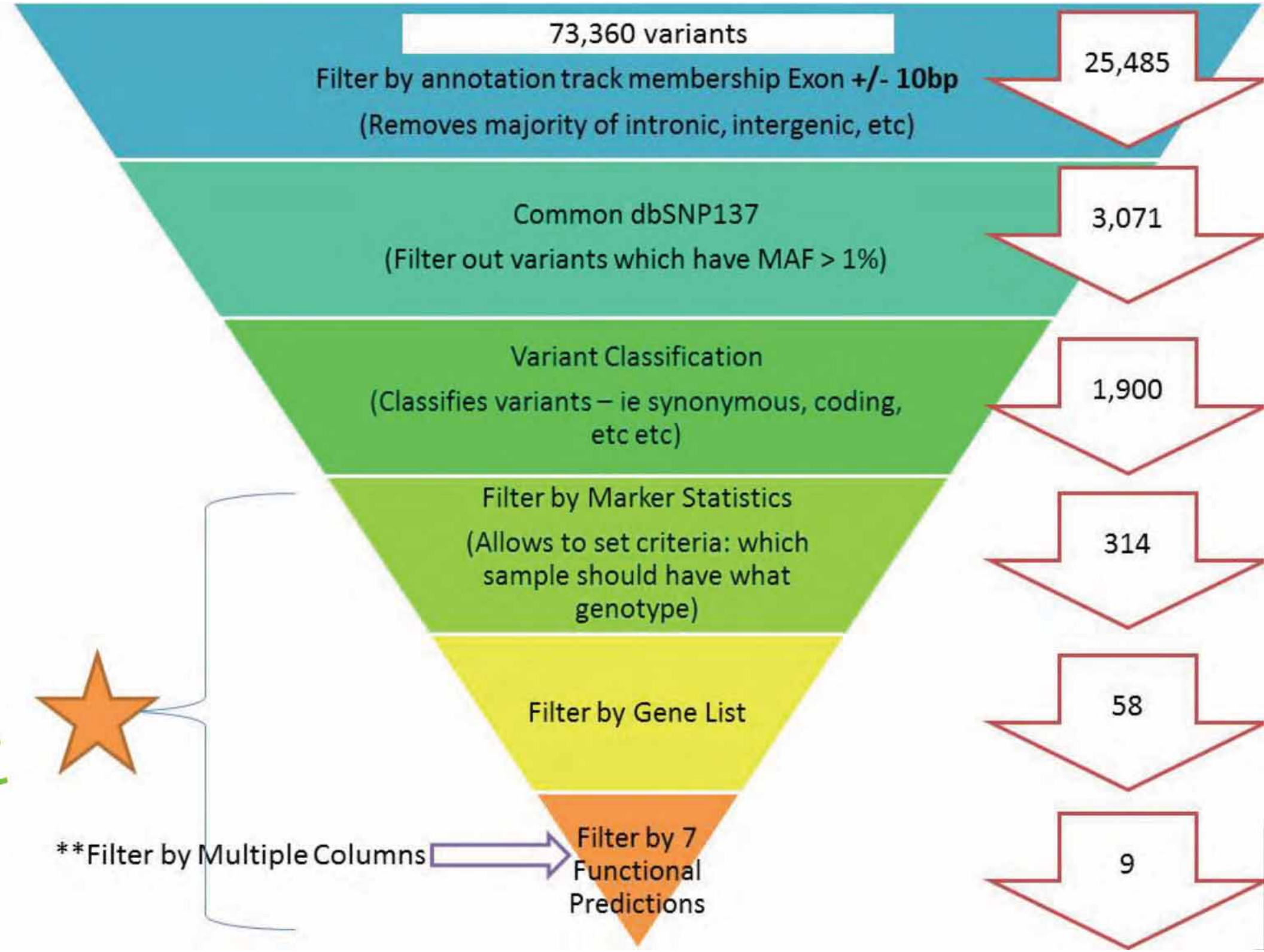
Filtering Assumptions

- Rare
- Autosomal Dominant model
- Coding-Nonsynonymous
- Conserved
- Biologically relevant

Filtering using SVS






73,360 variants
Filter by annotation track membership Exon +/- 10bp
(Removes majority of intronic, intergenic, etc)
25,485
Common dbSNP137
(Filter out variants which have MAF > 1%)
3,071
Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)
1,900
Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)
314
Filter by Gene List
58
**Filter by Multiple Columns
Filter by 7 Functional Predictions
9

Filter by Marker Statistics

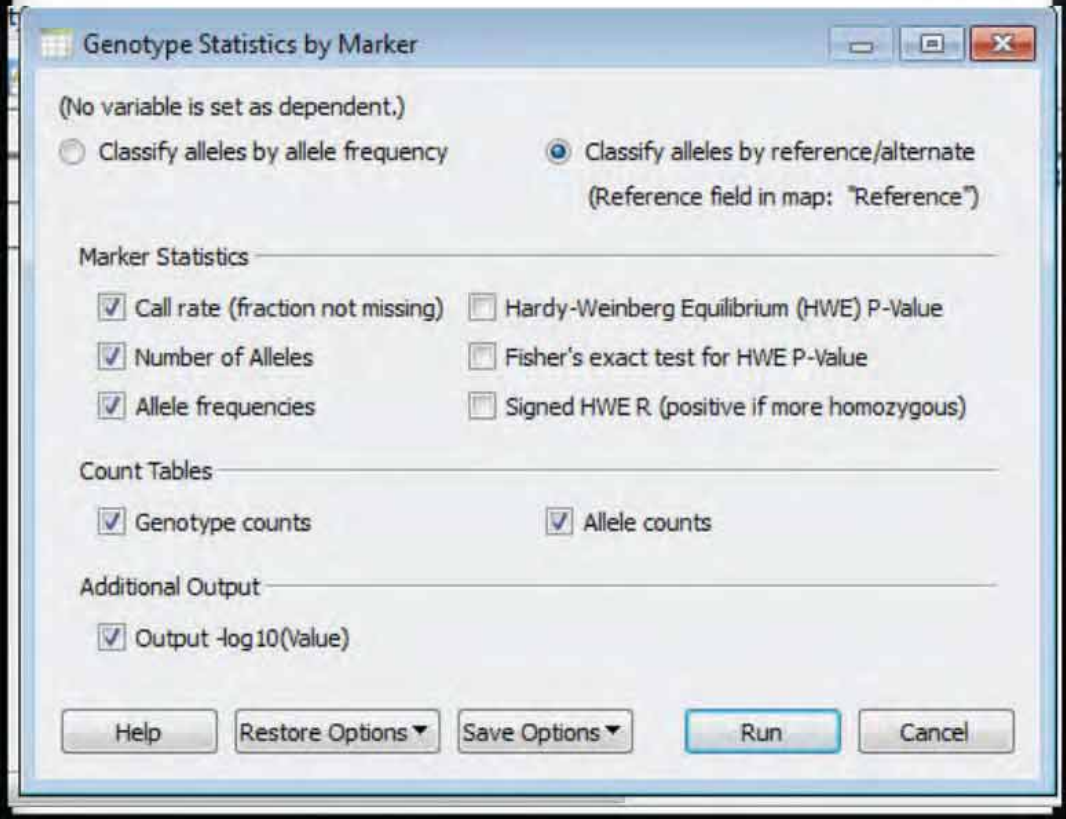
Use genotype information:

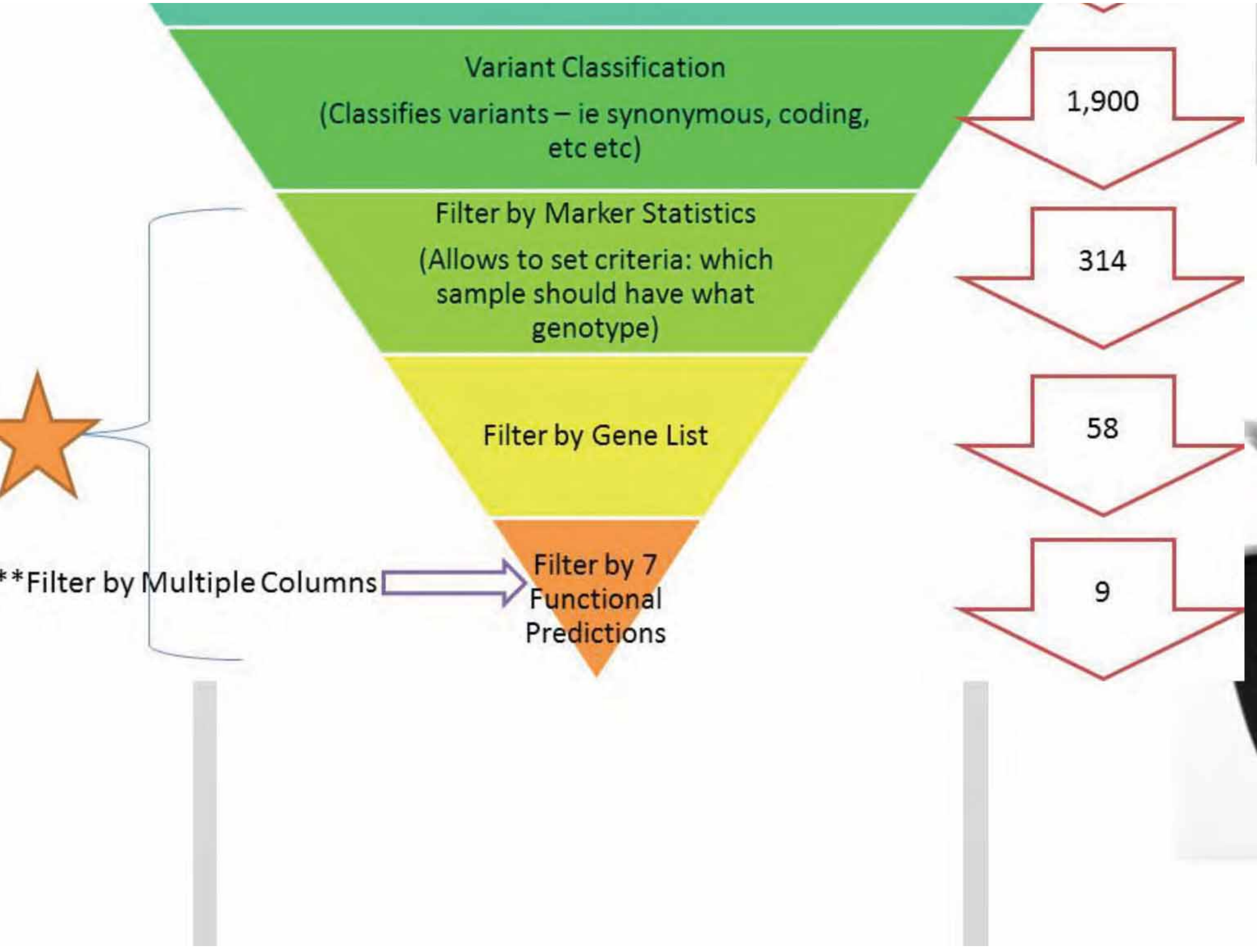
-Call rate

"X number of my samples have to have some sort of genotype call"

-Frequencies of genotype of interest

"X samples have/don't have ___ genotype calls"

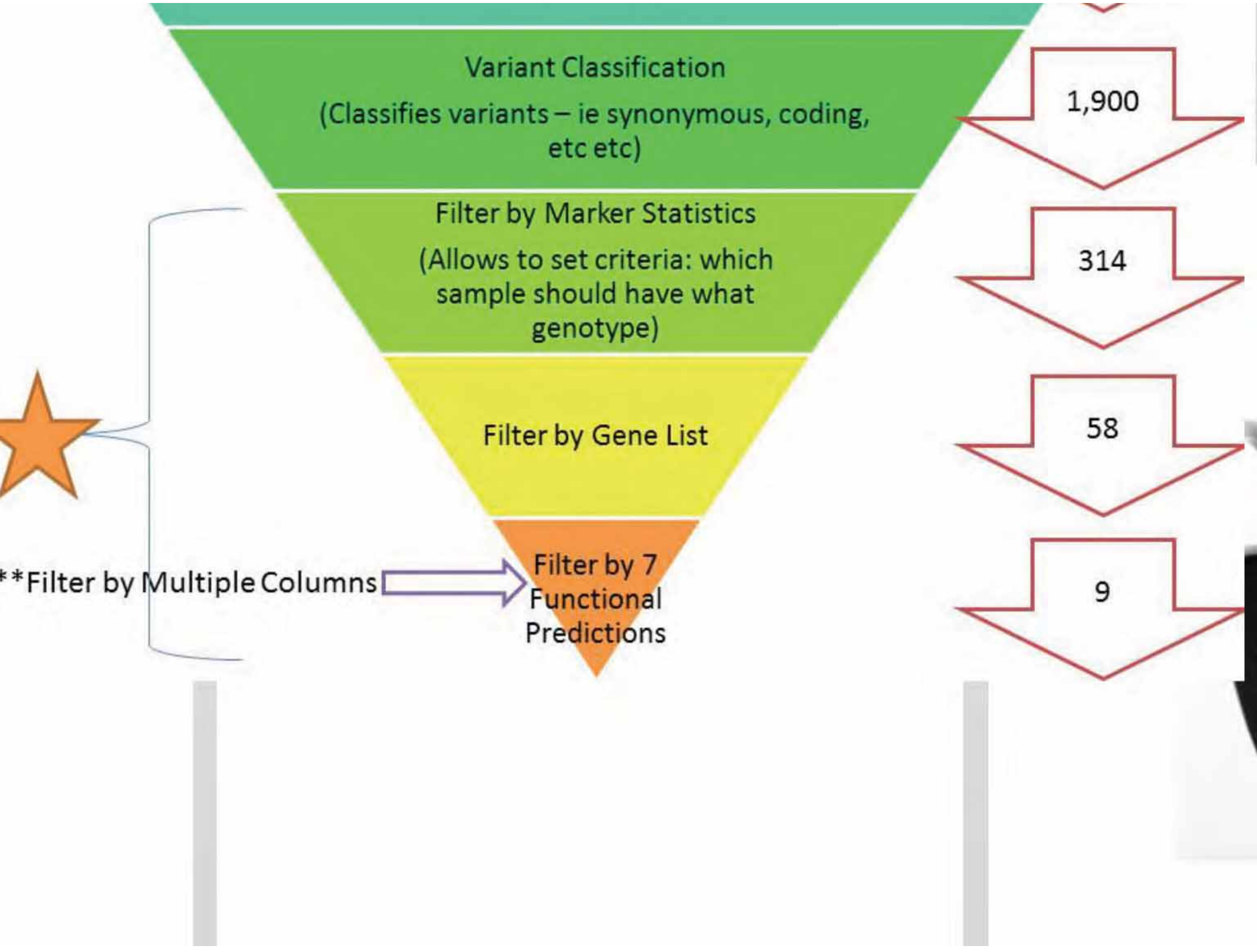



Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)
1,900
Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)
314
Filter by Gene List
58
**Filter by Multiple Columns
Filter by 7 Functional Predictions
9

Filter by gene list

Use list of candidate genes, associated genes, etc as a tool to filter

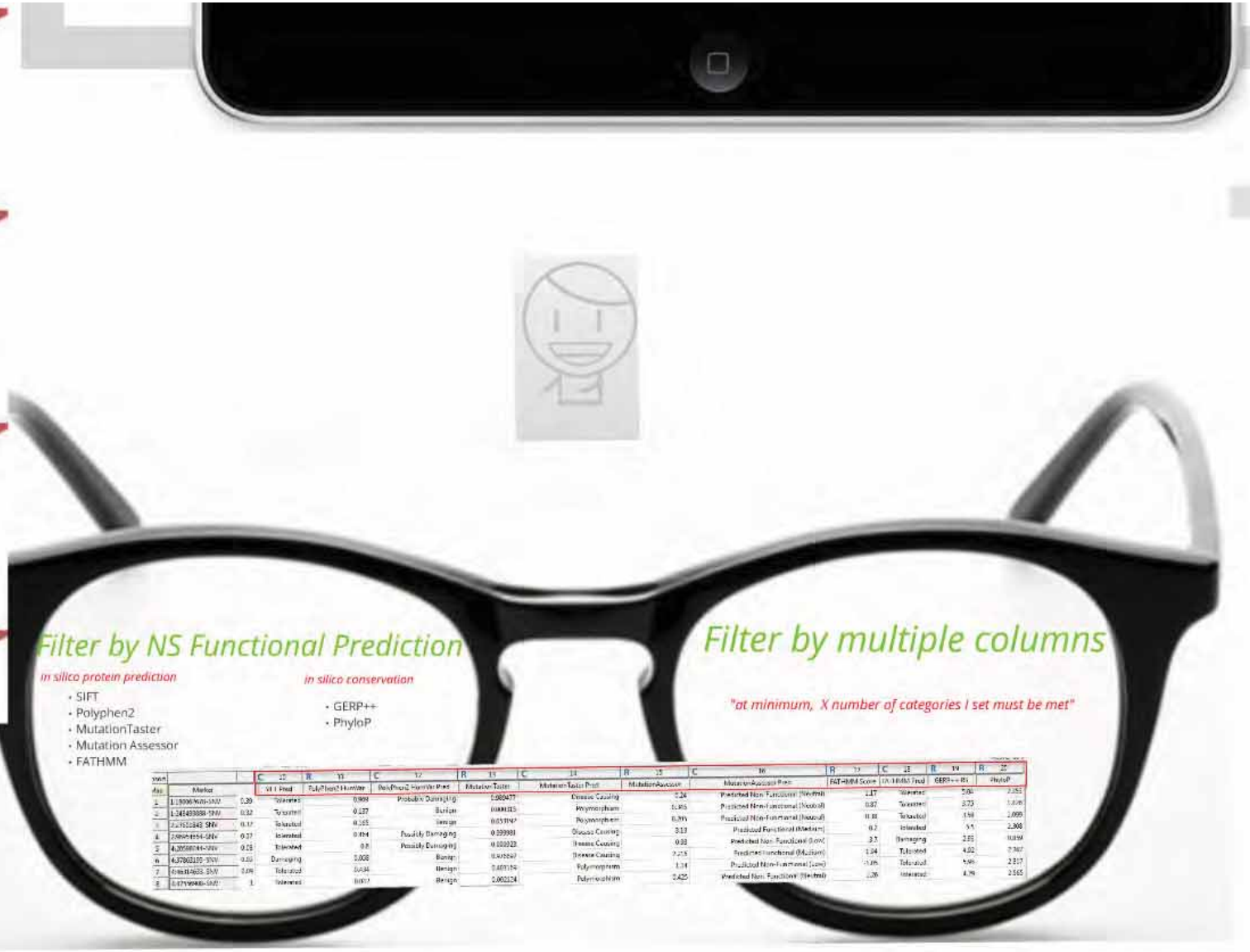
**Filter by Multi


Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)
1,900
Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)
314
Filter by Gene List
58
**Filter by Multiple Columns
Filter by 7 Functional Predictions
9

314

58

9


Filter by NS Functional Prediction
in silico protein prediction
• SIFT
• Polyphen2
• MutationTaster
• Mutation Assessor
• FATHMM
in silico conservation
• GERP++
• PhyloP
Filter by multiple columns
"at minimum, X number of categories I set must be met"

Filter by NS Functional Prediction

in silico protein prediction

- SIFT
- Polyphen2
- MutationTaster
- Mutation Assessor
- FATHMM

in silico conservation

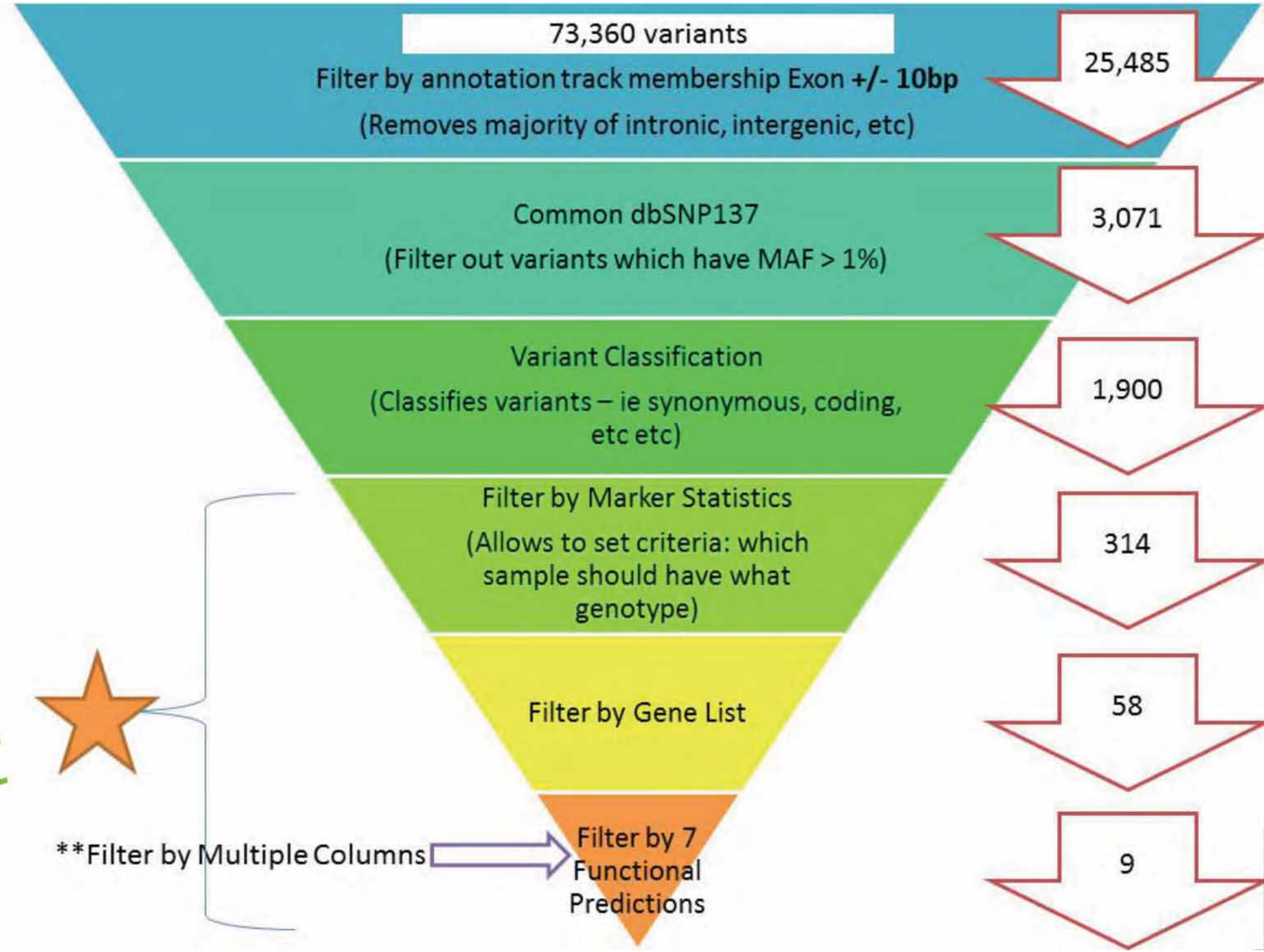
- GERP++
- PhyloP

nsort			C 10	R 11	C 12	R 13	C
Map	Marker		SIFT Pred	PolyPhen2 HumVar	PolyPhen2 HumVar Pred	MutationTaster	
1	1:186062678-SNV	0.39	Tolerated	0.969	Probably Damaging	0.989477	
2	1:243493888-SNV	0.32	Tolerated	0.137	Benign	0.000315	
3	2:27601843-SNV	0.32	Tolerated	0.165	Benign	0.053192	
4	2:96954854-SNV	0.07	Tolerated	0.464	Possibly Damaging	0.999981	
5	4:20598044-SNV	0.08	Tolerated	0.8	Possibly Damaging	0.999923	
6	4:37863193-SNV	0.02	Damaging	0.058	Benign	0.976697	
7	4:46314633-SNV	0.09	Tolerated	0.434	Benign	0.407169	
8	4:47556908-SNV	1	Tolerated	0.042	Benign	0.002124	

Filter by multiple columns

"at minimum, X number of categories I set must be met"

15	C 16	R 17	C 18	R 19	R 20
MutationAssessor	MutationAssessor Pred	FATHMM Score	FATHMM Pred	GERP++ RS	PhyloP
0.24	Predicted Non-Functional (Neutral)	-1.17	Tolerated	5.04	2.353
0.345	Predicted Non-Functional (Neutral)	0.87	Tolerated	3.75	1.428
0.205	Predicted Non-Functional (Neutral)	0.38	Tolerated	3.59	1.099
3.13	Predicted Functional (Medium)	0.2	Tolerated	5.5	2.308
0.93	Predicted Non-Functional (Low)	-3.7	Damaging	2.93	0.859
2.215	Predicted Functional (Medium)	1.94	Tolerated	4.92	2.367
1.14	Predicted Non-Functional (Low)	-1.05	Tolerated	5.95	2.817
0.425	Predicted Non-Functional (Neutral)	1.26	Tolerated	4.29	2.565


73,360 variants
Filter by annotation track membership Exon +/- 10bp
(Removes majority of intronic, intergenic, etc)
25,485
Common dbSNP137
(Filter out variants which have MAF > 1%)
3,071
Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)
1,900
Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)
314
Filter by Gene List
58
**Filter by Multiple Columns
Filter by 7 Functional Predictions
9

Results

- A Novel nonsense mutation on exon 2 of COL2A1
- Not present in 2000 chromosomes nor in any public databases, highly conserved
- Exon 2 mutations is predominantly ocular-only phenotype
- Currently in submission

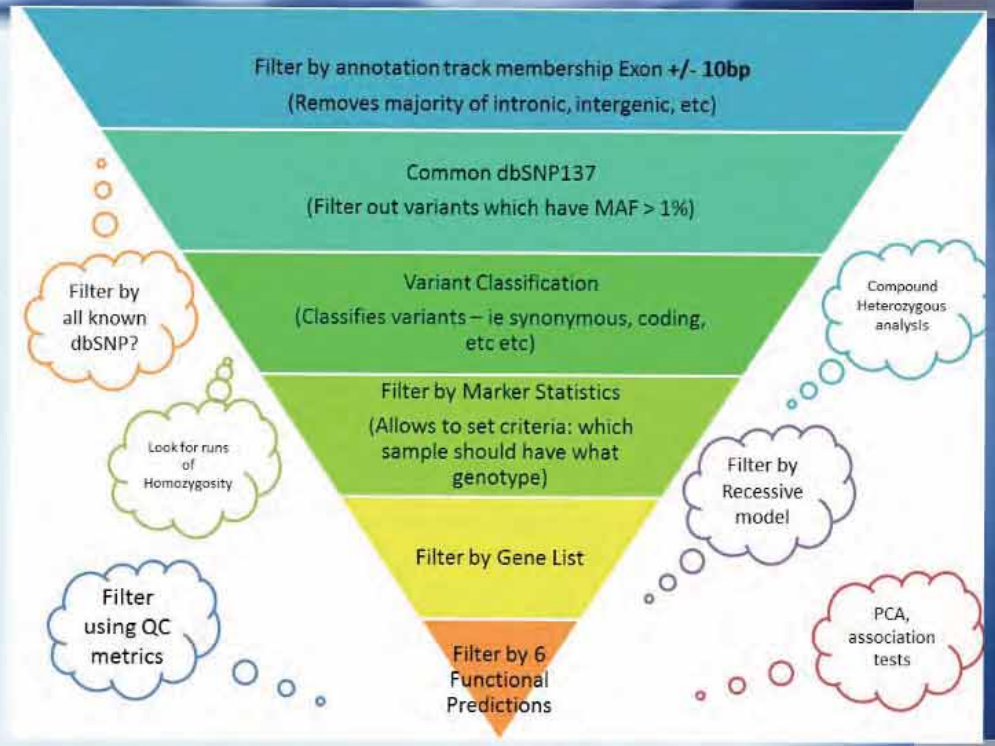
Challenges and Opportunities

-Filtering strategies will depend on your phenotype

-Not what's right or wrong...but think "efficiency"

-Filter to minimize re-analysis

-Understanding the study design



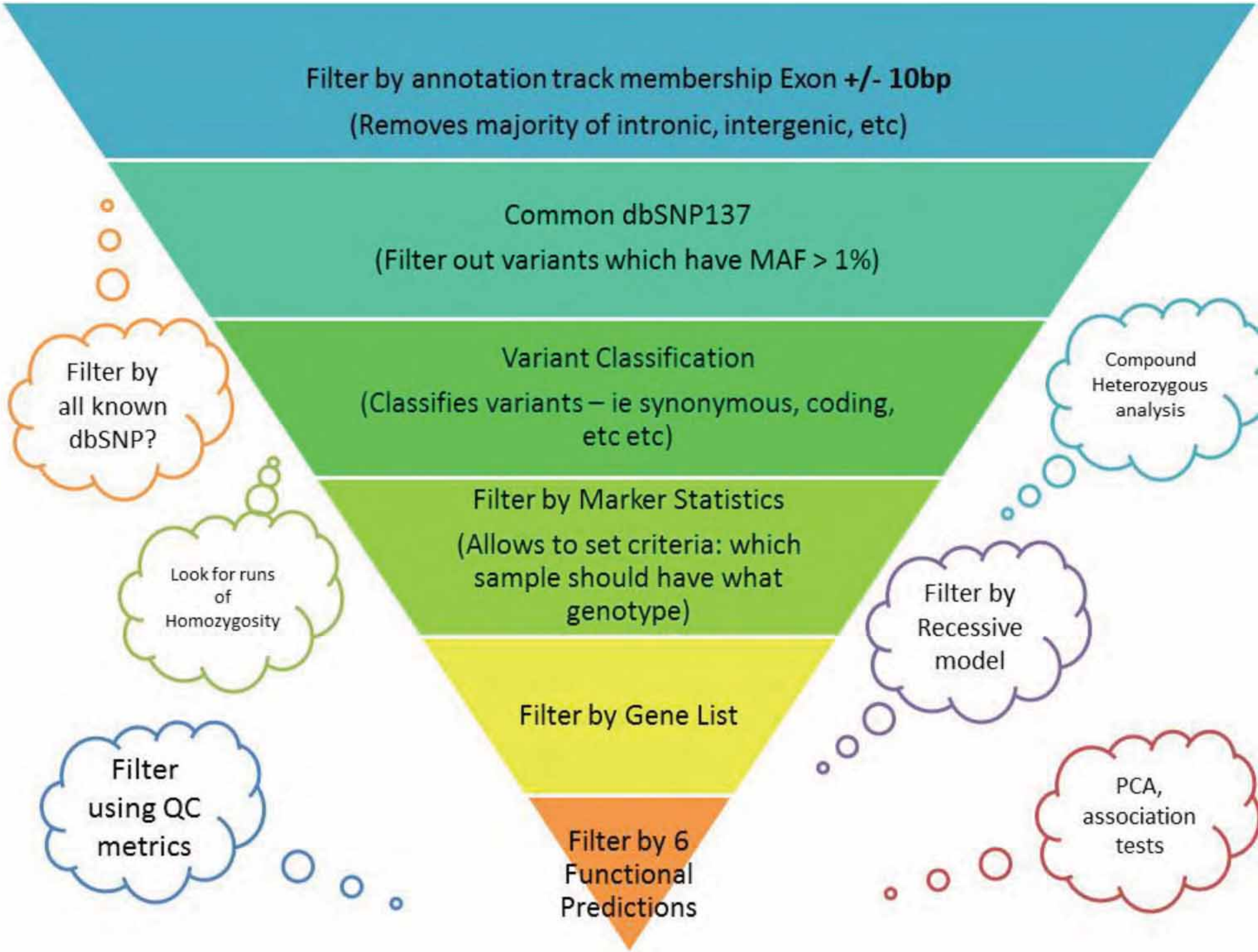
Using Golden Helix SVS

Pros:

- Customer service

Cons:

- Learning curve

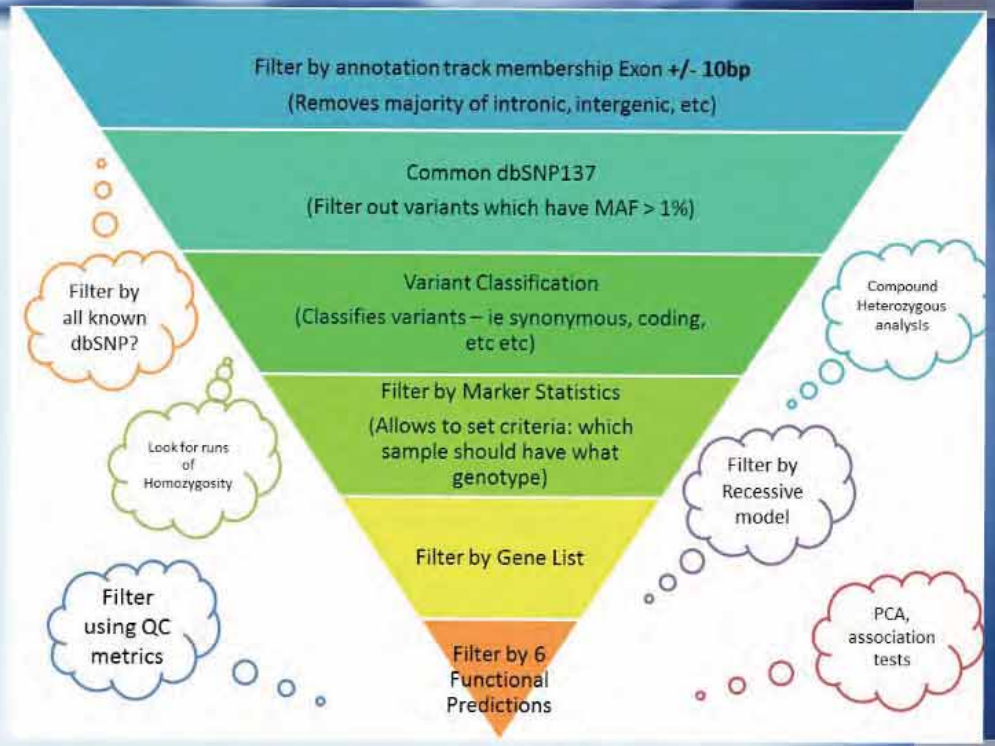

Filter by annotation track membership Exon +/- 10bp
(Removes majority of intronic, intergenic, etc)
Common dbSNP137
(Filter out variants which have MAF > 1%)
Variant Classification
(Classifies variants – ie synonymous, coding, etc etc)
Filter by Marker Statistics
(Allows to set criteria: which sample should have what genotype)
Filter by Gene List
Filter by 6 Functional Predictions
Filter by all known dbSNP?
Look for runs of Homozygosity
Filter using QC metrics
Compound Heterozygous analysis
Filter by Recessive model
PCA, association tests

-Filtering strategies will depend on your phenotype

-Not what's right or wrong...but think "efficiency"

-Filter to minimize re-analysis

-Understanding the study design



Using Golden Helix SVS

Pros:

- Customer service

Cons:

- Learning curve

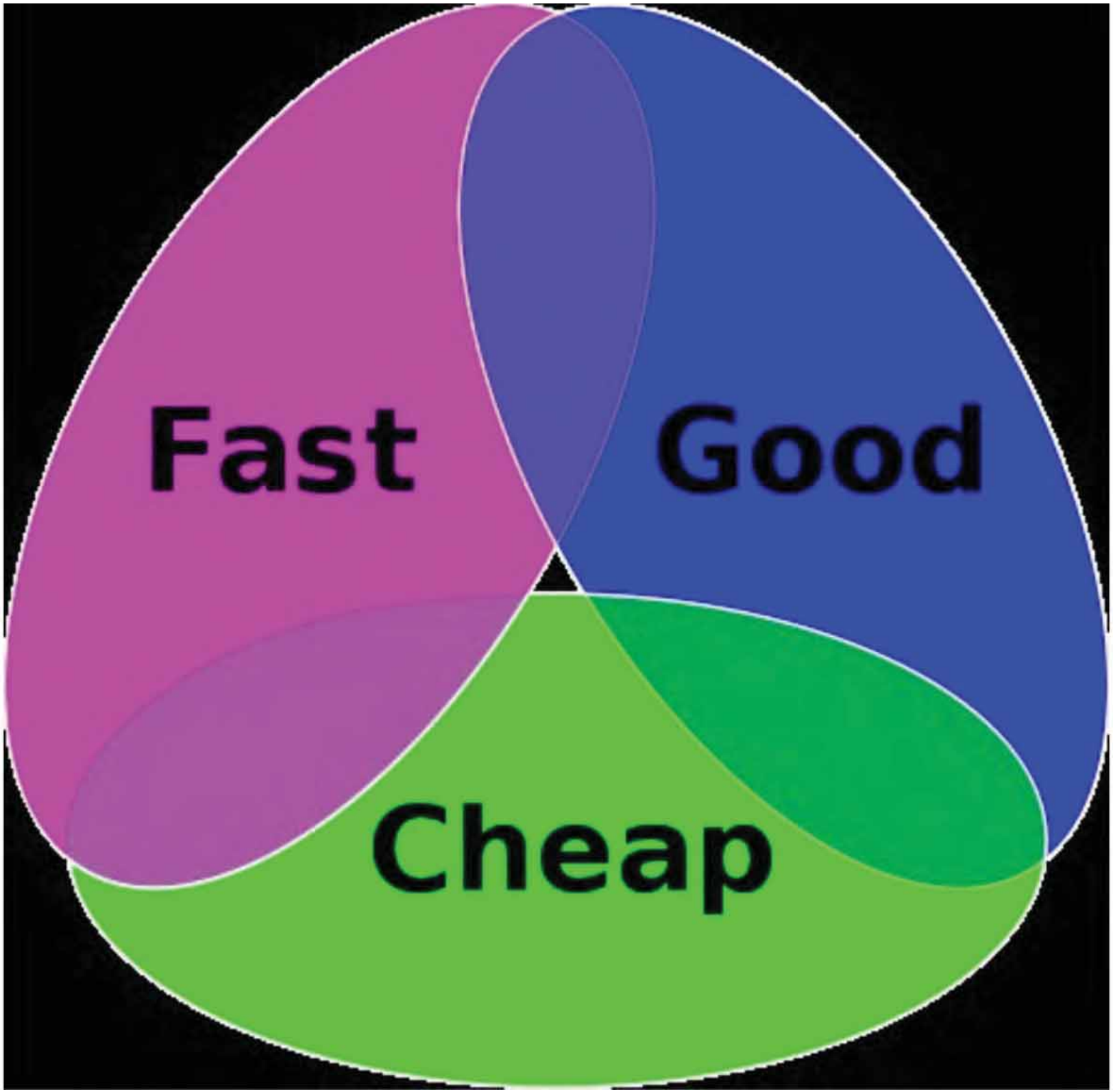
Using Golden Helix SVS

Pros:

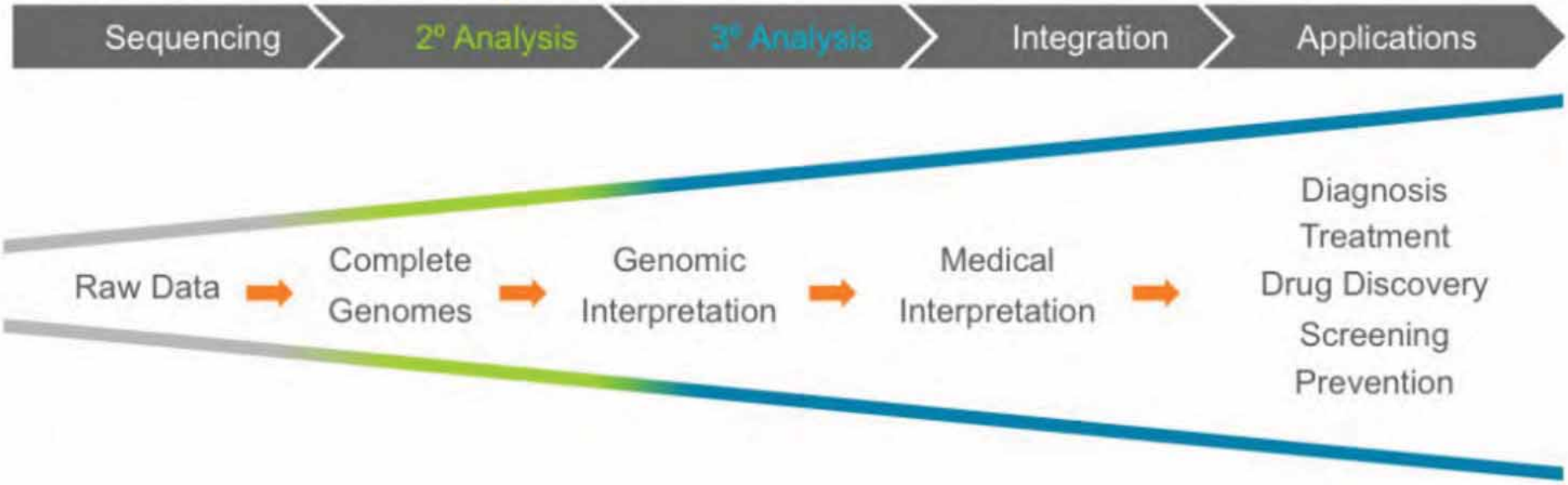
- Customer service
- Opportunties for everyone to participate
- Documentation/Internal QC
- Endless possibilities

Cons:

- Learning curve
- Cost prohibitive
- Continual updates


Fast
Good
Cheap

Genomics Landscape



http://venturebeat.com/2013/0

Acknowledgements

Center for Human Genetics

Terri Young, MD, MBA
Erica Nading, CGC
Diana Abbott, PhD
Yi-Ju Li, PhD
Xiaoyan Luo, PhD
Vincent Soler, MD
Caldwell Powell
Elizabeth St.Germain
Kristina Nicholson
Jenny Wei
Krystina Quow

Duke-NUS

Steve Rozen, PhD
John McPherson, PhD
Thomas Klemm
Stuart Tompson, PhD

Golden Helix

Greta Peterson, PhD
Autumn Laughbaum

A*STAR Singapore

Sebastian Maurer-Stroh, PhD
Vachiranee Limviphuvadh, PhD

Collaborators

Max Johnson, MD
Charles Johnson
Benjamin Bakall, MD
Ed Stone, MD
Barathi Veluchamy, PhD
Ravikanth Metlapally, PhD
Tammy Yanovitch, MD

Hudson Alpha

Shawn Levy, PhD
Braden Boone, PhD
Jack Wimbish

Questions?

Contact: Khanh-Nhat Tran-Viet
K.tranviet@duke.edu